PharmaSUG 2019 - Paper DS-239

More Traceability: Clarity in ADaM Metadata and Beyond

Wayne Zhong, Accretion Softworks
Richann Watson, DataRich Consulting
Daphne Ewing, CSL Behring
Jasmine Zhang, Boehringer Ingelheim

ABSTRACT

One of the fundamental principles of ADaM is that datasets and associated metadata must include traceability to facilitate the understanding of the relationships between analysis results, ADaM datasets, and SDTM datasets. The existing ADaM documents contain isolated elements of traceability, such as including SDTM sequence numbers, creating new records to capture derived analysis values, and providing excerpts of define.xml documentation.

An ADaM sub-team is currently developing a Traceability Examples Document with the goal of bringing these separate elements of traceability together and demonstrate how they function in detailed and complete examples. The examples cover a wide variety of practical scenarios; some expand on content from other CDISC documents, while others are developed specifically for the Traceability Examples Document. As members of the Traceability Examples ADaM sub-team, we are including in this PharmaSUG paper a selection of examples to show how traceability can bring transparency and clarity to your analyses.

INTRODUCTION

Clinical studies are conducted to test the safety and effectiveness of new drugs and therapies. Data collected from study participants is analyzed, and the results submitted to regulatory agencies and released to the public. To ensure the results are robust and verifiable, the steps by which the collected data is processed into the analysis results should be clearly documented. This is known as **Traceability**.

Why is this important? Suppose a new vaccine is showing promise, as seen in **Table 1** below.

Table 1: Sample Efficacy Table

Table xx.x Primary Efficacy Endpoint

ITT Population

	Drug n (%) (N=8000)	Control n (%) (N=8000)	Odds Ratio	P-Value
Occurrence of Primary Study Disease at 2 Years	8 (0.1%)	64 (0.8%)	0.1241	< 0.0001

This result table shows the study drug reduced the occurrence rate of disease by >80% over the control, a clear improvement. Due to the importance of this analysis in the submission, a reviewer decides to double check the computation of these values. The define.xml analysis results metadata (ARM) in **Table 2** shows one type of metadata the reviewer may find useful.

Table 2: Sample ARM

Display	Table xx.x Primary Efficacy Endpoint
Analysis Result	Occurrence of Primary Study Disease at 2 Years
Analysis Parameter(s)	PARAMCD = "PRI" (Primary Efficacy Endpoint)
Analysis Variable(s)	AVAL (Analysis Value)
Analysis Reason	SPECIFIED IN SAP
Analysis Purpose	PRIMARY OUTCOME MEASURE
Data References (incl. Selection Criteria)	ADEF [PARAMCD = "PRI" and ITTFL = "Y"]
Documentation	SAP Section 4.1
Programming Statements	[SAS Version 9.2] proc freq data=adef(where=(ittfl='Y' and paramcd='PRI')); table trt01pn*aval; exact or; run;

The metadata identifies the dataset used as ADEF along with the necessary subset conditions. The SAS® code snippet allows quick verification of table values using submitted data. This is an example of **metadata traceability**. **Table 3** below shows a slice of the dataset ADEF.

Table 3: Sample ADEF Records

USUBJID	SRCDOM	SRCSEQ	PARAMCD	PARAM	AVAL	AVALC
XYZ-01-001	PF	2	PRI	Primary Efficacy Endpoint	0	DISEASE
XYZ-01-002	LB	52	PRI	Primary Efficacy Endpoint	0	DISEASE
XYZ-01-003			PRI	Primary Efficacy Endpoint	1	NO DISEASE
XYZ-01-004			PRI	Primary Efficacy Endpoint	1	NO DISEASE

After verifying that the odds ratio and p-value are calculated correctly using the provided ADaM dataset, the reviewer continues to verify that the parameter itself is derived correctly. To facilitate this part of the review, **Table 4** provides variable metadata for the ADEF dataset.

Table 4: Sample ADEF Variable Level Metadata

Name	Variable Label	Variable Metadata
USUBJID	Unique Subject Identifier	ADSL.USUBJID
SRCDOM	Source Data	If AVAL=0, identify whether the corresponding record is from PF or LB SDTM domain
SRCSEQ	Source Sequence	If AVAL=0, copy over the corresponding PFSEQ or LBSEQ value from the corresponding record
PARAMCD	Parameter Code	Set to "PRI"
PARAM	Parameter	Set to " Primary Efficacy Endpoint"

Name	Variable Label	Variable Metadata
AVAL	Analysis Value	If subject has a biopsy record in PF where PFTEST="BIOMARKER 1" and PFSTRESC="PRESENT" then set AVAL=0. Else if subject does not have any biopsy records in PF and has an enzyme record in LB where LBTEST="ENZYME A" and LBSTRESC="POSITIVE" then set AVAL=0. (note: if a biopsy absent record is present, do not check enzyme test records) Otherwise set AVAL=1 Refer to SAP section 4.1 for more details
AVALC	Analysis Value (C)	If AVAL=0 then set AVALC="DISEASE" If AVAL=1 then set AVALC="NO DISEASE"

Table 4 provides the derivation of the primary efficacy parameter in the variable AVAL, as another example of **metadata traceability**. In addition, when study disease is identified for a subject, the variables SRCDOM and SRCSEQ identify the exact record in SDTM that led to this determination. This is an example of **data point traceability**. With this, the data lineage from the efficacy table to SDTM is complete.

In the case that the traceability between analysis results and SDTM is incomplete, the reviewer may have to decrypt the code in submitted analysis programs or hold review question and answers cycles with the sponsor, both of which take time. Providing both metadata and data point traceability enables quick and efficient reviews of analysis data and results and thus is a cornerstone of a quality submission.

This paper provides five examples being developed as part of the ADaM Traceability Examples document. To fully understand this paper, it is expected that the reader be fluent in the CDISC standards ADaM, SDTM, and Define-XML. Documents about these and other CDISC standards can be downloaded from the CDISC website https://www.cdisc.org/. See the Recommended Reading section for a list of CDISC standards documents most pertinent to this paper.

Please note all examples (including data structures, algorithms, data flows, table shells) are for illustration purposes and are not meant to represent a standard way of analyzing data. The approach taken by the reader will be sponsor specific.

EXAMPLE 1: GENERAL OCCDS TRACEABILITY

Occurrence analysis is the counting of subjects with a given record or term, and often includes a structured hierarchy of dictionary coding categories. Examples of data that fit into this structure include those used for typical analysis of Adverse Events, Concomitant Medications, and Medical History. The structure for occurrence analysis dataset is usually one record per record in the corresponding SDTM domain.

DATA FLOW

Figure 1 illustrates the data flow for the general ADAE, which can be created by merging ADaM.ADSL and SDTM.AE.

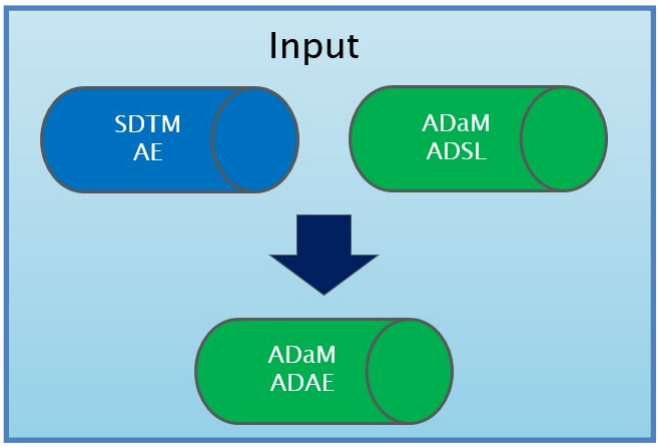


Figure 1: Example of General OCCDS Traceability Data Flow

TRACEABLITY NEEDS

Table 5 below shows the dataset metadata for ADAE, it describes the structure of the dataset as being one record per record from the corresponding SDTM AE domain.

Table 5: Dataset Metadata for ADAE

Dataset Name	Dataset Description	Dataset Structure	Class of Dataset
ADAE	Adverse Events Analysis Dataset	One record per record in SDTM domain AE (USUBJID AETERM ASTDT AENDT AESEQ).	OCCURRENCE DATA STRUCTURE

Table 6 below shows the variable metadata for ADAE. AE variables needed for analysis, such as AEDECOD, AEBODSYS, etc. are included in ADAE. In addition, variable AESEQ is retained to provide data point traceability on a record-to-record basis between AE and ADAE. Variables such as AESEV and AESTDTC may not be used directly in the analysis, but they supplement the derivation of ASEV, ASTDT, ASTDTF by enabling a quick review of the derivation of these variables. For the same reason, TRTSDT, TRTEDT are copied over from ADSL to support the derivation of variable TRTEMFL. In addition, variables such as SEX, RACE are copied over from ADSL to support analysis.

Please note only a small subset of variables are kept in this example for presentation purposes, a full ADAE dataset will include more variables.

Table 6: Variable Metadata for ADAE

Name	Variable Label	Variable Metadata
STUDYID	Study Identifier	AE.STUDYID
USUBJID	Unique Subject Identifier	AE.USUBJID
AESEQ	Sequence Number	AE.AESEQ
AETERM	Reported Term for the Adverse Event	AE.AETERM
AEDECOD	Dictionary-Derived Term	AE.AEDECOD (MedDRA Version 11.1)
AEBODSYS	Body System or Organ Class	AE.AEBODSYS
TRTEMFL	Treatment Emergent Analysis Flag	If ADSL.TRTSDT <= ASTDT<=(ADSL.TRTEDT +14) then TRTEMFL='Y'

Name	Variable Label	Variable Metadata
AESTDTC	Start Date/Time of Adverse Event	AE.AESTDTC
AESEV	Severity/Intensity	AE.AESEV
ASEV	Analysis Severity/Intensity	If AE.AESEV='MILD' then ASEV='Mild' Else if AE.AESEV='MODERATE' then ASEV='Moderate' Else if AE.AESEV is equal to 'SEVERE' or Severity/Intensity is missing then ASEV='Severe'
ASTDT	Analysis Start Date	Numeric version by converting AE.AESTDTC from character ISO8601 format to SAS format, applying imputation rules as specified in the SAP or metadata.
ASTDTF	Analysis Start Date Imputation Flag	If start date is completely missing or missing the year then ASTDTF ='Y' Else if start date has month missing then ASTDTF='M' Else if start date has day missing then ASTDTF='D'
TRTSDT	Date of First Exposure to Treatment	ADSL.TRTSDT
TRTEDT	Date of Last Exposure to Treatment	ADSL.TRTEDT
SEX	Sex	ADSL.SEX
RACE	Race	ADSL.RACE

INPUT AND ANALYSIS DATA

The dataset in **Table 7** below shows a sample implementation of the ADAE metadata above.

Table 7: ADAE Sample Records

Row	STUDYID	USUBJID	AESEQ	AETERM	AEDECOD	AEBODSYS
1	XYZ	XYZ-001-001	1	HEADACHE	Headache	Nervous system disorders
2	XYZ	XYZ-001-001	2	CHRONIC BACK PAIN	Back pain	Musculoskeletal and connective tissue disorders
3	XYZ	XYZ-001-001	3	NOSE BLEEDING RIGHT NOSTRIL	Epistaxis	Respiratory, thoracic and mediastinal disorders
4	XYZ	XYZ-001-001	4	PROBLEMS OF HYPOTENSION	Hypotension	Vascular disorders
5	XYZ	XYZ-001-001	5	HEADACHE	Headache	Nervous system disorders

Row	AESEV	ASEV	AESTDTC	ASTDT	ASTDTF
1 (cont)	MILD	Mild	2006-01	01JAN2006	D
2 (cont)	MODERATE	Moderate	2006-01-21	21JAN2006	
3 (cont)	MILD	Mild	2006-01-22	22JAN2006	
4 (cont)	MILD	Mild		23JAN2006	Y
5 (cont)		Severe	2006-01-24	24JAN2006	

Row	TRTEMFL	TRTSDT	TRTEDT	SEX	RACE
1 (cont)		23JAN2006	15MAY2006	M	ASIAN
2 (cont)		23JAN2006	15MAY2006	M	ASIAN

Row	TRTEMFL	TRTSDT	TRTEDT	SEX	RACE
3 (cont)		23JAN2006	15MAY2006	M	ASIAN
4 (cont)	Y	23JAN2006	15MAY2006	M	ASIAN
5 (cont)	Y	23JAN2006	15MAY2006	M	ASIAN

OTHER USES

This OCCDS ADAE example demonstrated data point traceability to source SDTM records. It also supported derived variables in ADAE by keeping source variables used in the derivation from AE and ADSL.

This process can be generalized to any OCCDS ADaM dataset. The dataset will include variables needed for analysis, variables supporting derivations, and variables providing data point traceability. The metadata provides the description of the dataset and derivations can be verified easily by using this approach.

EXAMPLE 2: USING AN INTERMEDIATE DATASET FOR BDS TRACEABILITY

Often the analysis that needs to be conducted is not based on a simple straightforward endpoint. It typically involves some complex derivation of the analysis endpoint(s). When dealing with complex derivations it may be beneficial to capture all the necessary data that is associated with the derivation for a patient in one dataset. These data can be an assessment date, protocol violation, prohibited medication or other intervention. For the purpose of this example these data will be collectively referred to as events. The creation of an intermediate data can aid in the reviews as well as the understanding of the analysis endpoint(s). The Therapeutic Area Data Standards User Guide for Breast Cancer section 5.3 includes an example of an intermediate dataset that supports both the time-to-event dataset and the best response dataset. Content of this example is being expanded in the ADaM Traceability Examples document to illustrate how to maintain traceability when there is an intermediate dataset.

DATA FLOW

Data for the intermediate dataset (ADEVENT) contains all events which are used in the derivation of two different datasets that will be used for analysis, time-to-event (ADTTE) and best overall response (ADRESP). Both ADTTE and ADRESP datasets are based off the RECIST assessment; however, ADTTE is also based off additional events. These additional events needed to determine time-to-event are date of randomization, date of the use of a prohibited medication and date of study disposition. **Figure 2** illustrates the various datasets that can be used in the creation of the intermediate dataset, ADEVENT.

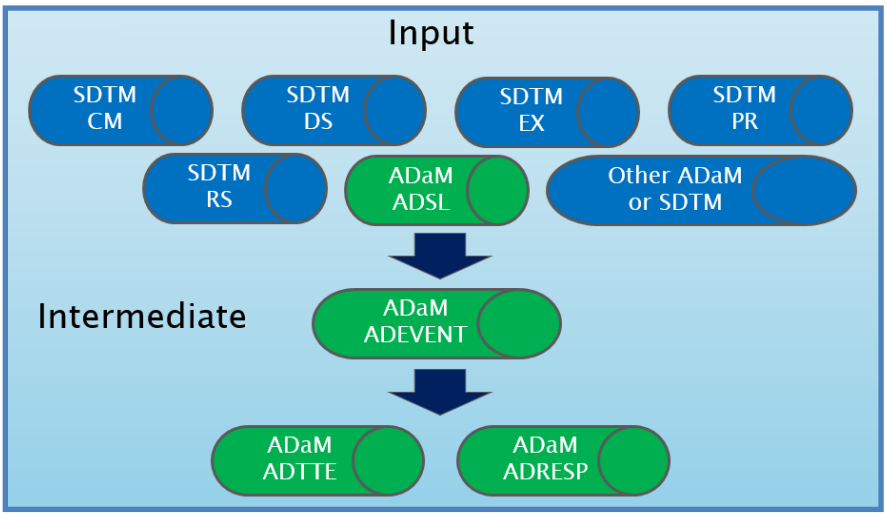


Figure 2: Example of Using an Intermediate Dataset for Traceability Data Flow

Note that 'Other SDTM or ADaM' datasets in this example data flow pertain to any dataset that will contain an event that is necessary for the derivation of the analysis endpoint(s). This is not an implication that all SDTM or ADaM data needs to be incorporated into one dataset. Only the data used for the derivations should be included.

TRACEABLITY NEEDS

Keeping all events (e.g., disposition events, response assessments, concomitant medications) provides the necessary data to derive the analysis endpoint(s) and is part of the traceability for both analysis datasets, ADTTE and ADRESP. **Table 8** shows the variable metadata for the intermediate dataset (ADEVENT) while **Table 9** and **Table 10** show the variable metadata for the analysis datasets (ADTTE and ADRESP). **Table 11** shows the Parameter-Value-Level metadata for select variables for both the intermediate dataset and the two analysis datasets.

Note that only variables used to illustrate the concept of traceability are shown in the tables.

Table 8: Variable Metadata for ADEVENT

Variable Name	Variable Label	Type	Codelist / Controlled Terms	Variable Metadata
STUDYID	Study Identifier	Char		xx.STUDYID that corresponds to input dataset used for USUBJID.
USUBJID	Unique Subject Identifier	Char		xx.USUBJID that corresponds to the input dataset used for the assignment of AVAL
ASEQ	Analysis Sequence Number	Num		Sequential number for associating a record number in the ADEVENT dataset. A unique number per subject, per parameter, per parameter qualifier, per analysis start date.
ASTDT	Analysis Start Date	Num		The date that the event occurred is the corresponding --DTC variable for each PARAMCD converted to numeric date format. RS.RSDTC when PARAMCD = 'ASSESS' DS.DSSTDTC when PARAMCD = 'DISPOSIT' AE.AESTDTC or MH.MHSTDTC or DV.DVSTDTC or CM.CMSTDTC or PR.PRSTDTC or some other source data for an event which prevents further assessments when PARAMCD = 'EVENT'.

Variable Name	Variable Label	Type	Codelist / Controlled Terms	Variable Metadata
ASTDY	Analysis Start Relative Day	Num		The number of days from randomization to the date of the reported event. ASTDT - ADSL.RANDDT + 1
PARQUAL	Parameter Qualifier	Char	INVESTIGATOR; CENTRAL	INVESTIGATOR for investigator-based tumor response assessments. CENTRAL for central imaging tumor response assessments. Otherwise set to null.
PARAMCD	Parameter Code	Char	ASSESS; DISPOSIT; TRTM; EVENT	If RECIST assessment, then PARAMCD = 'ASSESS' If disposition event, then PARAMCD = 'DISPOSIT' If study treatment, then PARAMCD = 'TRTM' If event that is a protocol violation or prevents further assessments, then PARAMCD = 'EVENT'
AVALC	Analysis Value (C)	Char		Reported Assessment associated with the ASTDT.
SRCDOM	Source Data	Char		***See parameter-level metadata below***
SRCVAR	Source Variable	Char		***See parameter-level metadata below***
SRCSEQ	Source Sequence Number	Num		The sequence number --SEQ or ASEQ of the row in the dataset identified in the SRCDOM that relates to the analysis value being derived.
ANL01FL	Analysis Flag 01	Char	Y	Identifies whether the event can be used in time-to-event analysis. If assessment is prior to baseline or after a censoring event, then they are not included.

Table 9: Variable Metadata for ADTTE

Variable Name	Variable Label	Type	Codelist / Controlled Terms	Variable Metadata
STUDYID	Study Identifier	Char		ADEVENT.STUDYID
USUBJID	Unique Subject Identifier	Char		ADEVENT.USUBJID
PARQUAL	Parameter Qualifier	Char	INVESTIGATOR; CENTRAL	For PARAMCD = 'PFS', INVESTIGATOR for investigator-based assessments; CENTRAL for central imaging assessments. Otherwise set to null.
PARAMCD	Parameter Code	Char	PFS; OS; DOR	RS.RSDTC when PARAMCD = 'ASSESS' DS.DSSTDTC when PARAMCD = 'DISPOSIT' AE.AESTDTC or MH.MHSTDTC or DV.DVSTDTC or CM.CMSTDTC, or PR.PRSTDTC or some other source data for an event which prevents further assessments when PARAMCD = 'EVENT'. If duration of response, then set to 'DOR.'
AVAL	Analysis Value	Num		The numeric value representing the time from the reference start date to the date of the analysis. (Where an analysis may be an assessment, disposition, or event recorded as documented in ADEVENT.) ***See parameter-level metadata below***
CNSR	Censored	Num	1; 0	Censoring Value. Set to 1 if value is censored based on rules. ***See parameter-level metadata below***

Variable Name	Variable Label	Type	Codelist / Controlled Terms	Variable Metadata
EVNTDESC	Event or Censoring Description	Char	DOCUMENTED PROGRESSION; DEATH; DISEASED PROGRESSED; CENSORED AT TIME OF LAST ASSESSMENT	***See parameter-level metadata below***
SRCDOM	Source Data	Char	ADEVENT	SRCDOM = 'ADEVENT'
SRCVAR	Source Variable	Char	ASTDY	If PARAMCD = 'DOR', then SRCVAR is null. Otherwise SRCVAR = 'ASTDY.'
SRCSEQ	Source Sequence Number	Num		If PARAMCD = 'DOR', then SRCSEQ is null. Otherwise SRCSEQ = ADEVENT.ASEQ

Table 10: Variable Metadata for ADRESP

Variable Name	Variable Label	Type	Codelist / Controlled Terms	Variable Metadata
STUDYID	Study Identifier	Char		ADEVENT.STUDYID
USUBJID	Unique Subject Identifier	Char		ADEVENT.USUBJID
PARQUAL	Parameter Qualifier	Char	INVESTIGATOR; CENTRAL	INVESTIGATOR for investigator-based assessments. CENTRAL for central imaging assessments.
PARAMCD	Parameter Code	Char	BOR	RS.RSDTC when PARAMCD = 'ASSESS'
AVAL	Analysis Value	Num	1; 2; 3; 4	DS.DSSTDTC when PARAMCD = 'DISPOSIT' AE.AESTDTC or MH.MHSTDTC or DV.DVSTDTC or CM.CMSTDTC, or PR.PRSTDTC or some other source data for an event which prevents further assessments when PARAMCD = 'EVENT'. ***See parameter-level metadata below***
AVALC	Analysis Value (C)	Char	CR; PR; SD; PD	***See parameter-level metadata below***
SRCDOM	Source Data	Char	ADEVENT	SRCDOM = 'ADEVENT'
SRCVAR	Source Variable	Char	AVALC	SRCVAR = 'AVALC'
SRCSEQ	Source Sequence Number	Num		ADEVENT.ASEQ

Table 11: Value Level Metadata

Dataset	Variable Name	Where	Type	Derivation / Comment
ADEVENT	SRCDOM	PARAMCD = 'ASSESS'	Char	Set to 'RS'
ADEVENT	SRCDOM	PARAMCD = 'DISPOSIT'	Char	Set to 'DS'
ADEVENT	SRCDOM	PARAMCD = 'TRTM'	Char	Set to 'EX'
ADEVENT	SRCDOM	PARAMCD = 'EVENT'	Char	Set to SDTM domain where the event that is a protocol violation or prevents further assessment is represented
ADEVENT	SRCVAR	PARAMCD = 'ASSESS'	Char	Set to 'AVALC'
ADEVENT	SRCVAR	PARAMCD = 'DISPOSIT'	Char	Set to 'DSDECOD'
ADEVENT	SRCVAR	PARAMCD = 'TRTM'	Char	Set to 'EXTRT'

Dataset	Variable Name	Where	Type	Derivation / Comment
ADEVENT	SRCVAR	PARAMCD = 'EVENT'	Char	Set to SDTM variable where the event that is a protocol violation or prevents further assessment is represented
ADTTE	AVAL	PARAMCD = 'PFS'	Num	ADEVENT.ASTDY when ADEVENT.ANL01FL = 'Y and ADEVENT.PARAMCD = 'ASSESS' and disease progressed.
ADTTE	AVAL	PARAMCD = 'OS'	Num	ADEVENT.ASTDY when ADEVENT.ANL01FL = 'Y' and ADEVENT.PARAMCD = 'EVENT' and ADEVENT.AVALC = 'DEATH' or when ADEVENT.ANL01FL = 'Y' and ADEVENT.PARAMCD = 'DISPOSIT' and maximum ADEVENT.ASTDY
ADTTE	AVAL	PARAMCD = 'DOR'	Num	Time from best response (CR or PR) to when disease progressed
ADTTE	CNSR	PARAMCD = 'PFS'	Num	If disease did not progress, then CNSR = 1. Otherwise CNSR = 0.
ADTTE	CNSR	PARAMCD = 'OS'	Num	If subject did not die, then CNSR = 1. Otherwise CNSR = 0.
ADTTE	CNSR	PARAMCD = 'DOR'	Num	If after having best response disease did not progress, then CNSR = 1. Otherwise CNSR = 0.
ADTTE	EVNTDESC	PARAMCD = 'PFS'	Char	If CNSR = 0, then EVNTDESC = 'DOCUMENTED PROGRESSION'. Otherwise, if CNSR = 1, then EVNTDESC = 'CENSORED AT TIME OF LAST ASSESSMENT'.
ADTTE	EVNTDESC	PARAMCD = 'OS'	Char	If CNSR = 0, then EVNTDESC = 'DEATH'. Otherwise, if CNSR = 1, then EVNTDESC = 'CENSORED AT TIME OF LAST ASSESSMENT'.
ADTTE	EVNTDESC	PARAMCD = 'DOR'	Char	If CNSR = 0, then EVNTDESC = 'DISEASED PROGRESSED'. Otherwise, if CNSR = 1, then EVNTDESC = 'CENSORED AT TIME OF LAST ASSESSMENT'.
ADRESP	AVAL	PARAMCD = 'BOR'	Num	If AVALC = 'CR' then AVAL = 1 If AVALC = 'PR' then AVAL = 2 If AVALC = 'SD' then AVAL = 3 If AVALC = 'PD' then AVAL = 4
ADRESP	AVALC	PARAMCD = 'BOR'	Char	ADEVENT.ANL01FL = 'Y' and ADEVENT.PARAMCD = 'ASSESS,' then set to the best response where ranking from best to worst is 'CR,' 'PR,' 'SD,' 'PD'.

Use of the Provisional Variable

Some ADaM examples make use of a special purpose parameter-qualifying variable, PARQUAL. PARQUAL is a proposed variable with restricted defined usage currently that is under consideration by the CDISC ADaM team. Note, however, that this variable is incompatible with the current ADaMIG, which states, "PARAM must include all descriptive and qualifying information relevant to the analysis purpose of the parameter."

One advantage of using this qualifying variable would be that the same controlled terminology could be used for PARAM and PARAMCD. If ratified, the use of this variable would be limited and would only apply to situations where there is no change in the interpretation of AVAL. The example above utilizes this special purpose variable, but readers should fully appreciate that this variable is provisional and is subject to change.

INPUT AND ANALYSIS DATA

The data flow diagram and the variable level metadata show that there are several data sources that contain various events that are used in the determination of the analysis endpoint(s). **Tables 12 – 17** contain example input data which is used to illustrate how ADEVENT is created based on the variable metadata in **Table 8**. Only variables pertinent to the example are included here.

Table 12: Input Data Example CM

Row	STUDYID	USUBJID	CMSEQ	CMTRT	CMSTDTC
1	ABC	ABC-001	1	TAMOXIFEN	2014-03-31

Table 13: Input Data Example DS

Row	STUDYID	USUBJID	DSSEQ	DSTERM	DSDECOD	DSSTDTC
1	ABC	ABC-001	1	INFORMED CONSENT	INFORMED CONSENT OBTAINED	2013-12-03
2	ABC	ABC-001	2	RANDOMIZED	RANDOMIZED	2013-12-29
3	ABC	ABC-002	1	INFORMED CONSENT	INFORMED CONSENT OBTAINED	2013-10-25
4	ABC	ABC-002	2	RANDOMIZED	RANDOMIZED	2013-11-10

Table 14: Input Data Example EX

Row	STUDYID	USUBJID	EXSEQ	EXTRT	EXSTDTC	EXENDTC
1	ABC	ABC-001	1	DRUG A	2014-01-01	2014-01-01
2	ABC	ABC-001	2	DRUG A	2014-03-30	2014-03-30
3	ABC	ABC-002	1	DRUG B	2013-11-13	2013-11-13
4	ABC	ABC-002	2	DRUG B	2013-12-29	2013-12-29

Table 15: Input Data Example PR

Row	STUDYID	USUBJID	PRSEQ	PRTRT	PRSTDTC
1	ABC	ABC-002	1	LUMPECTOMY	2013-12-14

Table 16: Input Data Example RS

Row	STUDYID	USUBJID	RSSEQ	RSTESTCD	RSSTRESC	RSEVAL	RSDTC
1	ABC	ABC-001	1	OVRLRESP	PD	INVESTIGATOR	2013-12-30
2	ABC	ABC-001	2	OVRLRESP	SD	CENTRAL	2013-12-31
3	ABC	ABC-001	3	OVRLRESP	SD	INVESTIGATOR	2014-01-21
4	ABC	ABC-001	4	OVRLRESP	SD	CENTRAL	2014-01-22
5	ABC	ABC-001	5	OVRLRESP	PR	INVESTIGATOR	2014-02-13
6	ABC	ABC-001	6	OVRLRESP	PR	CENTRAL	2014-02-14
7	ABC	ABC-001	7	OVRLRESP	PR	INVESTIGATOR	2014-03-06
8	ABC	ABC-001	8	OVRLRESP	PR	CENTRAL	2014-03-07
9	ABC	ABC-001	9	OVRLRESP	PD	INVESTIGATOR	2014-03-28

Row	STUDYID	USUBJID	RSSEQ	RSTESTCD	RSSTRESC	RSEVAL	RSDTC
10	ABC	ABC-001	10	OVRLRESP	PD	CENTRAL	2014-03-29
11	ABC	ABC-002	1	OVRLRESP	PD	INVESTIGATOR	2013-11-11
12	ABC	ABC-002	2	OVRLRESP	PD	CENTRAL	2013-11-12
13	ABC	ABC-002	3	OVRLRESP	SD	INVESTIGATOR	2013-12-01
14	ABC	ABC-002	4	OVRLRESP	SD	CENTRAL	2013-12-02
15	ABC	ABC-002	5	OVRLRESP	PR	INVESTIGATOR	2013-12-27
16	ABC	ABC-002	6	OVRLRESP	PR	CENTRAL	2013-12-28

Table 17: Input Data for ADSL

Row	STUDYID	USUBJID	TRT01P	TRT01SDT	TR01EDT	RANDFL	RANDDT
1	ABC	ABC-001	DRUG A	2014-01-01	2014-03-30	Y	2013-12-29
2	ABC	ABC-002	DRUG B	2013-11-13	2013-12-29	Y	2013-11-10

Utilizing the SRCDOM, SRCVAR, and SRCSEQ variables in ADEVENT allows us to trace each event back to the original source dataset. For example, **Table 18** shows the resulting intermediate dataset ADEVENT. With ADEVENT available ADTTE and ADRESP can be created based on the variable level metadata in **Table 9** and **Table 10**, respectively. Additionally, SRCDOM, SRCVAR and SRCSEQ variables in ADTTE and ADRESP allows us to trace each endpoint back to the source record and variable in the ADEVENT dataset.

Table 19 shows the results of ADTTE and **Table 20** shows the results of ADRESP.

Table 18: Intermediate Data Example ADEVENT

Row	USUBJID	ASEQ	ASTDT	ASTDY	*PARQUAL*	PARAMCD	AVALC	ANL01FL	SRCDOM	SRCVAR	SRCSEQ
1	ABC-001	1	2013-12-29	-4		DISPOSIT	RANDOMIZED		DS	DSDECOD	2
2	ABC-001	2	2013-12-30	-2	*INVESTIGATOR*	ASSESS	PD	Y	RS	RSSTRESC	1
3	ABC-001	3	2013-12-31	-1	*CENTRAL*	ASSESS	SD	Y	RS	RSSTRESC	2
4	ABC-001	4	2014-01-01	1		TRTM	DRUG A	Y	EX	EXTRT	1
5	ABC-001	5	2014-01-21	20	*INVESTIGATOR*	ASSESS	SD	Y	RS	RSSTRESC	3
6	ABC-001	6	2014-01-22	22	*CENTRAL*	ASSESS	SD	Y	RS	RSSTRESC	4
7	ABC-001	7	2014-02-13	44	*INVESTIGATOR*	ASSESS	PR	Y	RS	RSSTRESC	5
8	ABC-001	8	2014-02-14	45	*CENTRAL*	ASSESS	PR	Y	RS	RSSTRESC	6
9	ABC-001	9	2014-03-06	65	*INVESTIGATOR*	ASSESS	PR	Y	RS	RSSTRESC	7
10	ABC-001	10	2014-03-07	66	*CENTRAL*	ASSESS	PR	Y	RS	RSSTRESC	8
11	ABC-001	11	2014-03-28	87	*INVESTIGATOR*	ASSESS	PD	Y	RS	RSSTRESC	9
12	ABC-001	12	2014-03-29	88	*CENTRAL*	ASSESS	PD	Y	RS	RSSTRESC	10
13	ABC-001	13	2014-03-30	89		TRTM	DRUG A	Y	EX	EXTRT	2
14	ABC-001	14	2014-03-31	90		EVENT	TAMOXIFEN		CM	CMTRT	1
15	ABC-002	1	2013-11-10	-3		DISPOSIT	RANDOMIZED		DS	DSDECOD	2
16	ABC-002	2	2013-11-11	-2	*INVESTIGATOR*	ASSESS	PD	Y	RS	RSSTRESC	1

Row	USUBJID	ASEQ	ASTDT	ASTDY	*PARQUAL*	PARAMCD	AVALC	ANL01FL	SRCDOM	SRCVAR	SRCSEQ
17	ABC-002	3	2013-11-12	-1	*CENTRAL*	ASSESS	PD	Y	RS	RSSTRESC	2
18	ABC-002	4	2013-11-13	1		TRTM	DRUG B	Y	EX	EXTRT	1
19	ABC-002	5	2013-12-01	19	*INVESTIGATOR*	ASSESS	SD	Y	RS	RSSTRESC	3
20	ABC-002	6	2013-12-02	20	*CENTRAL*	ASSESS	SD	Y	RS	RSSTRESC	4
21	ABC-002	7	2013-12-14	32		EVENT	LUMPECTOMY		PR	PRTRT	1
22	ABC-002	8	2013-12-27	45	*INVESTIGATOR*	ASSESS	PR		RS	RSSTRESC	5
23	ABC-002	9	2013-12-28	46	*CENTRAL*	ASSESS	PR		RS	RSSTRESC	6
24	ABC-002	10	2013-12-29	47		TRTM	DRUG B		EX	EXTRT	2

Table 19: Output Data Example ADTTE

Row	STUDYID	USUBJID	*PARQUAL*	PARAMCD	AVAL	CNSR	EVNTDESC	SRCDOM	SRCVAR	SRCSEQ
1	ABC	ABC-001	*INVESTIGATOR*	PFS	87	0	DOCUMENTED PROGRESSION	ADEVENT	ASTDY	11
2	ABC	ABC-001	*CENTRAL*	PFS	88	0	DOCUMENTED PROGRESSION	ADEVENT	ASTDY	12
3	ABC	ABC-001		OS	89	1	CENSORED AT TIME OF LAST ASSESSMENT	ADEVENT	ASTDY	13
4	ABC	ABC-001		DOR	44	0	DISEASED PROGRESSED			
5	ABC	ABC-002	*INVESTIGATOR*	PFS	19	1	CENSORED AT TIME OF LAST ASSESSMENT	ADEVENT	ASTDY	5
6	ABC	ABC-002	*CENTRAL*	PFS	20	1	CENSORED AT TIME OF LAST ASSESSMENT	ADEVENT	ASTDY	6
7	ABC	ABC-002		OS	1	1	CENSORED AT TIME OF LAST ASSESSMENT	ADEVENT	ASTDY	4

Table 20: Output Data Example ADRESP

Row	STUDYID	USUBJID	*PARQUAL*	PARAMCD	AVAL	AVALC	SRCDOM	SRCVAR	SRCSEQ
1	ABC	ABC-001	INVESTIGATOR	BOR	2	PR	ADEVENT	AVALC	7
2	ABC	ABC-001	CENTRAL	BOR	2	PR	ADEVENT	AVALC	8
3	ABC	ABC-002	INVESTIGATOR	BOR	3	SD	ADEVENT	AVALC	5
4	ABC	ABC-002	CENTRAL	BOR	4	SD	ADEVENT	AVALC	6

With the use of an intermediate dataset, it is easy to see why a specific value was used for an endpoint or why a subject was censored. For example, if we were to examine patient ABC-002 during a review process. We see that the RS data shows the subject had a Partial Response (**Table 16** rows 15 and 16). However, in ADTTE for PFS the subject is censored (**Table 19** rows 5 and 6) at the last assessment indicating the subject did not have a partial response. Also, in ADRESP the subject shows a best overall response of stable disease (**Table 20** rows 3 and 4). Why the discrepancy in the data? With the use of the intermediate dataset ADEVENT it contains all the events that were considered during the determination of the analysis endpoint and it is evident the patient had a LUMPECTOMY which is a prohibited medical procedure. (**Table 18** row 21). Therefore, any assessments after the procedure would not be eligible for the determination of the endpoint.

OTHER USES

The intermediate dataset will allow the reviewer to consider the impact on the analysis if an alternate value is selected. In addition, it will allow for the analysis endpoint to use a different date and/or value to perform sensitivity analyses.

EXAMPLE 3: TRACEABILITY WITH MULTIPLE INPUT DATASETS IN OCCDS

The unique --SEQ option from the OCCDS structure allows only one SDTM input for each OCCDS dataset. However, there are instances when it is necessary for multiple SDTM domains to be used as inputs to a single OCCDS dataset. The following example shows the scenario when rows from multiple input datasets are stacked together to form a single analysis dataset for occurrence analyses.

DATA FLOW

The analysis ADAECE combines both the spontaneous and pre-specified occurrences of the adverse event.

The SDTM AE domain is used as source of spontaneous adverse events, while the SDTM CE domain is used as source of pre-specified information about local infusion site reactions. The subject level ADSL dataset contains the subject-level data that could be used for the analysis, e.g. population flags, treatment arms and reference date information.

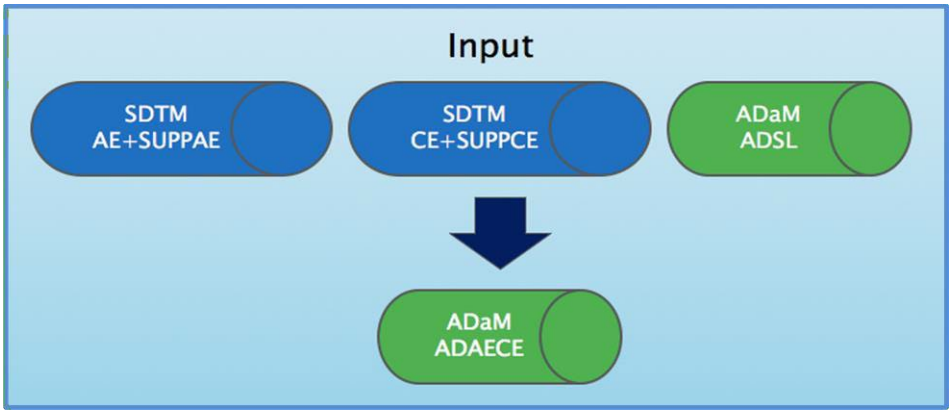


Figure 3: Example of Using Multiple Input Datasets in OCCDS for Traceability Data Flow

TRACEABLITY NEEDS

In this example, a separate eCRF page was provided to capture the pre-specified information about local infusion site reactions, while the spontaneous infusion reactions were captured on the standard adverse event eCRF. For the standard adverse event and infusion reactions summaries, the study statistician decided that both the pre-specified events that occurred and spontaneous events would be reported. The denominator for these displays is the Safety Population.

For this example, we are assuming a double-blind parallel design, Study Drug + Standard of Care versus Standard of Care.

Table 21: Example Table Showing Traceability Needs of Using Multiple Input Datasets in OCCDS

Table xx.x Summary of All Adverse Events by System Organ Class, MedDRA Preferred Term and Treatment Group

Safety Population

System Organ Class MedDRA Preferred Term	Study Drug + Standard of Care (N=XXX)	Standard of Care (N=XXX)
Subjects with at least one Adverse Event	xx (xx.x)	xx (xx.x)
Blood and lymphatic system disorders	xx (xx.x)	xx (xx.x)
Anaemia deficiencies	xx (xx.x)	xx (xx.x)
Lymphadenopathy	xx (xx.x)	xx (xx.x)

The “combined preferred term” in the structure showed in **Table 22** indicates the verbatim term can come from both the AE and CE source dataset.

Table 22: Dataset Metadata for ADAECE

Dataset	Description	Class	Structure	Keys
ADAECE	Adverse/Clinical Events Analysis Dataset	OCCURRENCE DATA STRUCTURE	One record per subject per combined preferred term per start datetime	STUDYID, USUBJID, UDECOD, ASTDT

Since multiple SDTM domains are stacked and used as input to the OCCDS ADAE dataset, SRCSEQ and SRCDOM are used to facilitate data point traceability instead of --SEQ. Using SRC* variables is a deviation from the current unique --SEQ option from the OCCDS structure defined in the ADaM OCCDS v1.0. Including the SRCSEQ and SRCDOM into the OCCDS dataset is under review within the ADaM subteam and further details will be released in OCCDS v1.1.

Variables with a U* prefix such as UTERM, UBODSYS, UDECOD, USTDTC and UENDTC are non-standard custom variables added to stack unmodified content from multiple existing SDTM domains with the same root variables into one field for analysis and traceability. For example, both CE and AE have the root variable BODSYS, Body System or Organ Class in the same dictionary version, and since AEBODSYS and CEBODSYS are stacked without modification for analysis, the variable UBODSYS was created to denote this. This variable extends beyond variables used for analysis and can be used for traceability variables such as USTDTC and UENDTC.

Table 23: Variable Metadata for ADAECE

Variable Name	Variable Label	Codelist / Controlled Terms	Variable Metadata
STUDYID	Study Identifier	XYZ	ADSL.STUDYID
USUBJID	Unique Subject Identifier		ADSL.USUBJID
SAFFL	Safety Population Flag	Y'='Yes' 'N'='No'	ADSL.SAFFL
TRTA	Actual Treatment	SOC+SD=Standard of Care + Study Drug SOC=Standard of Care	ADSL.TRT01A
TRTSDT	Date of First Exposure to Treatment	yymmdd10.	ADSL.TRTSDT

Variable Name	Variable Label	Codelist / Controlled Terms	Variable Metadata
TRTSDTM	Datetime of First Exposure to Treatment	datetime20.	ADSL.TRTSDTM
SRCDOM	Source Data		Set to 'AE' if record is from AE dataset. Set to 'CE' if record is from CE dataset.
SRCSEQ	Source Sequence Number		Set to AE.AESEQ if record is from AE dataset. Set to CE.CESEQ if record is from CE dataset.
ACAT1	Analysis Category 1	ADVERSE EVENTS LOCAL INFUSION SITE REACTIONS	If record is from AE then ACAT1='ADVERSE EVENTS' Else ACAT1='LOCAL INFUSION SITE REACTIONS'
UTERM	Reported Term		AE.AETERM if record is from AE dataset CE.CETERM if record is from CE dataset
UDECOD	Dictionary-Derived Term		AE.AEDECOD if record is from AE dataset CE.CEDECOD if record is from CE dataset
UBODSYS	Body System or Organ Class		AE.AEBODSYS if record is from AE dataset CE.CEBODSYS if record is from CE dataset
USTDTC	Start Date/Time of Event	ISO8601	AE.AESTDTC if record is from AE dataset CE.CEBODSYS if record is from CE dataset
UENDTC	End Date/Time of Event	ISO8601	AE.AEENDTC if record is from AE dataset CE.CEBODSYS if record is from CE dataset
ASTDT	Analysis Start Date	yymmdd10.	<Producer will insert derivation here> For example: Date part of AESTDTC. If full date is present, convert to numeric. If Day is missing but year and month correspond with treatment start year and month, then set day to the start day of treatment, otherwise assume the first of the month. If Day and Month are missing but Year corresponds with treatment start year, then set month and day to treatment start month and day, otherwise assume January 1st. If start date is completely missing, do not impute.
ASTDTF	Analysis Start Date Imputation Flag	DATEF	If start date has month missing, then ASTDTF='M' Else if start date has day missing, then ASTDTF='D'
ASTDTM	Analysis Start Date/Time	datetime20.	<Producer will insert derivation here> For example: Convert AESTDTC to a numeric datetime variable
AREL	Analysis Causality		If record is from AE, then AREL=AE.AEREL converted to proper case; Else if the record is from CE, then AREL= 'Definitely Related'
ARELGR1	Pooled Causality Group 1	Related Not Related	If AREL in ('Definitely Related' 'Possibly Related' 'Probably Related'), then ARELGR1='Related'. Else if AREL in ('Not Related' 'Unlikely Related') then ARELGR1='Not Related'.
CEPRESP	Clinical Event Pre-Specified	'Y'='Yes'	CE.CEPRESP

INPUT AND ANALYSIS DATA

The sample data for the input AE dataset includes the spontaneous occurrences. For pre-specified occurrences, only the events that occurred were kept. In this example, the CE events, which occurred were considered to be candidates for adverse events included in ADAECE. For example, Rows 2 and 3 in CE domain (**Table 25**) did not occur, so the two records were not included.

Table 24: Input Data Example AE

Row	USUBJID	AESEQ	AETERM	AEDECOD	AEBODSYS	AREL	AESTDTC	AEENDTC
1	XYZ-001	1	DIARRHEA	Diarrhea	Gastrointestinal disorders	POSSIBLY RELATED	2014-05-15	2014-05-16
1	XYZ-001	2	HEADACHE	Headache	Nervous system disorders	POSSIBLY RELATED	2014-02	2014-02-17
2	XYZ-001	3	LOW NEUTROPHILS	Neutropenias	Blood and lymphatic system disorders	POSSIBLY RELATED	2014-04-14	2014-06-12
4	XYZ-001	4	PNEUMONIA	Pneumonia	Infections and infestations	POSSIBLY RELATED	2014-05-13	2014-05-15

Table 25: Input Data Example CE

Row	USUBJID	CESEQ	CETERM	CEDECOD	CEBODSYS	CEPRESP	CEOCCUR	CESTDTC	CEENDTC
1	XYZ-001	1	SWELLING AT THE INFUSION SITE	Edema peripheral	General disorders and administration site conditions	Y	Y	2014-02-15	2014-02-15
2	XYZ-001	2	PAIN AT THE INFUSION SITE	Pain	General disorders and administration site conditions	Y	N		
3	XYZ-001	3	RASH AT THE INFUSION SITE	Rash	Skin and subcutaneous tissue disorders	Y	N		
4	XYZ-001	4	REDNESS AT THE INFUSION SITE	Skin erythema	Skin and subcutaneous tissue disorders	Y	Y	2014-02-15	2014-02-15

Since multiple SDTM domains are stacked together, SRCSEQ and SRCDOM are used to point to the domain and record where the data came from. SRCVAR is not used in OCCDS because users tend to point back to an entire observation or record of information containing a particular “term” or “treatment” and all of its qualifiers rather than a single analysis value (AVAL/AVALC) as seen in BDS.

As seen in this example, many variables are copied directly into the OCCDS structure from the SDTM domains without modification. With stacking, the U* variable allows users to preserve the copy feature while stacking the same type data into the same column. U* is intended to indicate “Unmodified”.

The pre-specified flag is copied in from the Clinical Events domain to support ad hoc analyses involving comparisons between pre-specified and spontaneous events. Adverse Events in the AE domain are spontaneously reported and not pre-specified so AEPRESP was not included in the AE domain. Due to this, the producer of the analysis dataset kept CEPRESP rather than creating UPRESP.

Any changes in type, casing, or imputation requires the creation of A* variables as seen with AREL and ARELGR1, respectively.

Note that ACAT1 is used to indicate the type of data but was not required for analysis.

Since this is just an example, not all variables needed for analysis have been included.

Table 26: Output Data Example ADAECE

Row	USUBJID	SAFFL	TRTA	TRTSDT	TRTSDTM	SRCDOM	SRCSEQ	ACAT1
1	XYZ-001	Y	SOC + SD	2014-02-15	2014-02-15T10:05	AE	1	ADVERSE EVENTS
2	XYZ-001	Y	SOC + SD	2014-02-15	2014-02-15T10:05	CE	1	LOCAL INFUSION SITE REACTIONS
3	XYZ-001	Y	SOC + SD	2014-02-15	2014-02-15T10:05	AE	2	ADVERSE EVENTS
4	XYZ-001	Y	SOC + SD	2014-02-15	2014-02-15T10:05	AE	3	ADVERSE EVENTS
5	XYZ-001	Y	SOC + SD	2014-02-15	2014-02-15T10:05	AE	4	ADVERSE EVENTS
6	XYZ-001	Y	SOC + SD	2014-02-15	2014-02-15T10:05	CE	4	LOCAL INFUSION SITE REACTIONS

Row	UTERM	UDECOD	UBODSYS	USTDTC	UENDTC	AREL	ARELGR1	CEPRESP
1 (cont)	DIARRHEA	Diarrhea	Gastrointestinal disorders	2014-05-15	2014-05-16	Possibly Related	Related	
2 (cont)	SWELLING AT THE INFUSION SITE	Edema peripheral	General disorders and administration site conditions	2014-02-15	2014-02-15	Definitely Related	Related	Y
3 (cont)	HEADACHE	Headache	Nervous system disorders	2014-02	2014-02-17	Possibly Related	Related	
4 (cont)	LOW NEUTROPHILS	Neutropenias	Blood and lymphatic system disorders	2014-04-14	2014-06-12	Possibly Related	Related	
5 (cont)	PNEUMONIA	Pneumonia	Infections and infestations	2014-05-13	2014-05-15	Possibly Related	Related	
6 (cont)	REDNESS AT THE INFUSION SITE	Skin erythema	Skin and subcutaneous tissue disorders	2014-02-15	2014-02-15	Definitely Related	Related	Y

OTHER USES

This example considers input domains of the same general observation class. In addition, the idea can be extended to input datasets of varying observation classes with additional harmonization. This implementation will be included in OCCDS Version 1.1, which is under review.

EXAMPLE 4: TRACEABILITY WHEN USING A LOOK-UP TABLE

When deriving an analysis dataset, most of the information needed is directly available in or derived from SDTM data. The most basic approach is to collect the full array of needed variables from the subject for each analysis need. Sometimes, one-to-one relationships may exist between variable values based on individual events or findings and the variables that help describe or qualify them. When this occurs, collecting each individual value becomes unnecessary and increases the risk to data integrity. Often times, one variable is all that needs to be collected. Using a look-up table to include related values can reduce the amount of collected data and data-entry errors.

DATA FLOW

Data that can be derived based on values from data collected on the CRF can be placed in a look-up table (e.g., PMEDLKUP). Sorting the look-up table and the SDTM data by the common variable(s) allows a merge that will provide the additional/associated values needed for analyses. In this example, we will be able to identify categories of prohibited medications that need to be presented.

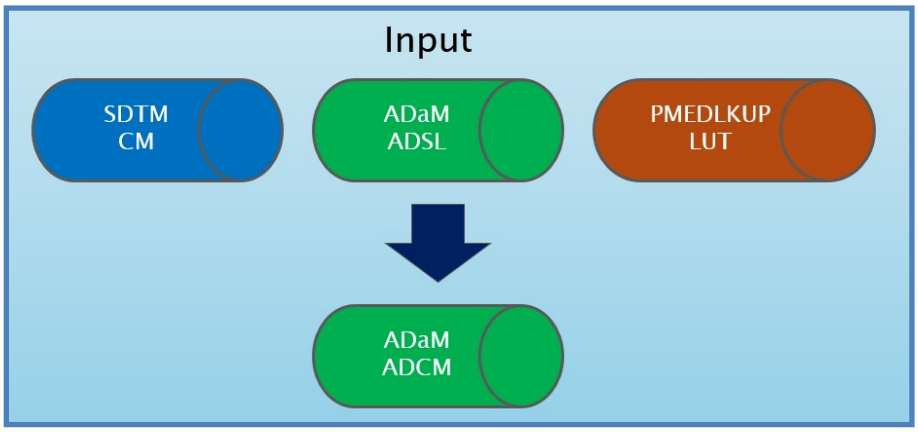


Figure 4: Example of Using a Look-Up Table for Traceability Data Flow

Note that the PMEDLKUP look-up table is used to generate a derived category value in the ADaM ADCM dataset. This provides the flexibility of defining categories outside of data collection to still be captured and analyzed as needed.

TRACEABLITY NEEDS

In this example, all medications were coded using WHODrug and Medication Name (CMDECOD) and Medication Class (CMCLASCD) were provided. There are several ATC Classes that are considered prohibited and these categories of prohibited medications need to be identified for summarization purposes.

Below is an example output that shows a count/percent of prohibited medication categories that are not clearly defined by an ATC Level.

Table 27: Example Table Showing Traceability Needs of Using Look-Up Tables

Table xx.x
Prohibited Medication During Study
Safety Population

Prohibited Medication Category/ Medication Name	Drug A (N=xx) n (%)	Drug B (N=xx) n (%)	Total (N=xx) n (%)
Category 1	xx (xxx.x)	xx (xxx.x)	xx (xxx.x)
Medication Name 1	xx (xxx.x)	xx (xxx.x)	xx (xxx.x)
Medication Name 2	xx (xxx.x)	xx (xxx.x)	xx (xxx.x)
Category 2	xx (xxx.x)	xx (xxx.x)	xx (xxx.x)
Medication Name 3	xx (xxx.x)	xx (xxx.x)	xx (xxx.x)
Medication Name 4	xx (xxx.x)	xx (xxx.x)	xx (xxx.x)

In order to categorize the data consistently, a look-up table will be used to perform the classification for the entire study. In the ADCM dataset, we want to include a variable that appropriately classifies each medication if it is one of the prohibited types. Below is the structure of the data we want to use to create the above summary:

Table 28: Variable Metadata for ADCM

Variable Name	Variable Label	Codelist / Controlled Terms	Variable Metadata
STUDYID	Study Identifier	XYZ	ADSL.STUDYID
DOMAIN	Domain Abbreviation	CM	CM.DOMAIN
USUBJID	Unique Subject Identifier		ADSL.USUBJID
CMSEQ	Sequence Number		CM.CMSEQ
CMDECOD	Standardized Medication Name	Drug Dictionary	CM.CMDECOD
CMCLASCD	Medication Class Code	Drug Dictionary	CM.CMCLASCD
ACAT1	Prohibited Medication Category		Derived: Populate by merging SDTM.CM with the look-up table, PMEDLKUP, by CMCLASCD. Set to the value of CATEGORY from the look-up table if available. Leave as null otherwise.

INPUT AND ANALYSIS DATA

The SDTM CM variables below contain the basic information that will be used for the above analysis.

Table 29: CM Input data

Row	STUDYID	DOMAIN	USUBJID	CMSEQ	CMDECOD	CMCLASCD
1	ABCD	CM	ABCD011001	1	METHYLPREDNISOLONE	H02AB
2	ABCD	CM	ABCD011001	2	ALPHARIX	J07BB
3	ABCD	CM	ABCD011001	3	BECLOMETHASONE	A07EA
4	ABCD	CM	ABCD011002	1	AMOXICILLINE	J01CA
5	ABCD	CM	ABCD011002	2	AZATHIOPRINUM	L04AX
6	ABCD	CM	ABCD021003	1	ADALIMUMAB	L04AB
7	ABCD	CM	ABCD021003	2	PREVENAR	J07AL

We will need to use the CMCLASCD provided by WHODrug for each medication to use as the look-up index to determine the prohibited medication category (ACAT1). Below is a sample look-up table for four categories.

Table 30: PMEDLKUP table identifying the relationship between ATC Class Codes and Categories

Category	CMCLASCD
Corticosteroid	A01AC
Corticosteroid	D07AC
Corticosteroid	H02AB
Corticosteroid	C05AA

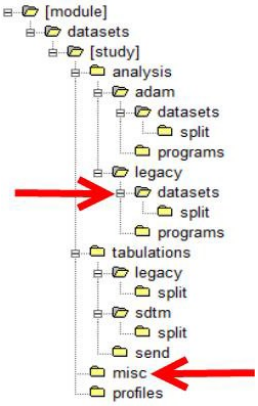
Category	CMCLASCD
Corticosteroid	A07EA
Corticosteroid	R01AD
Thiopurines	L01BB
Thiopurines	L04AX
Insulins	A10AD

By merging the SDTM CM data with this look-up table by the CMCLASCD value, we will be able to populate ACAT1 as defined in the ADCM metadata description.

Table 31: Output Data Example ADCM

Row	STUDYID	DOMAIN	USUBJID	CMSEQ	CMDECOD	CMCLASCD	ACAT1
1	ABCD	CM	ABCD011001	1	METHYLPREDNISOLONE	H02AB	Corticosteroid
2	ABCD	CM	ABCD011001	2	ALPHARIX	J07BB	
3	ABCD	CM	ABCD011003	3	BECLOMETHASONE	A07EA	
4	ABCD	CM	ABCD011002	1	AMOXICILLINE	J01CA	
5	ABCD	CM	ABCD011002	2	AZATHIOPRINUM	L04AX	Thiopurines
6	ABCD	CM	ABCD021003	1	ADALIMUMAB	L04AB	
7	ABCD	CM	ABCD021003	2	PREVENAR	J07AL	

When using look-up tables to generate values in ADaM datasets, there are options on how to submit the lookup table. The PMEDLKUP file can be submitted as an XPT file (in misc folder or legacy datasets folder):



Reference to the look-up table in the Analysis Data Reviewer's Guide (ADRG) should be documented as well. The actual look-up table can be included in the appendix of the ADRG and referenced in the section of the document that uses it.

OTHER USES

When data entry can include unique values needed for analysis, but other details related to what has been entered has a one-to-one relationship to another associate value, look-up tables can be a simple way to maintain high quality data that look-up tables can be used to create consistent PARAM values from PARAMCD values.

EXAMPLE 5: TRACEABILITY WHEN MULTIPLE ANALYSIS TIME POINTS ARE NEEDED ON THE SAME ROW

The BDS ADaM standard presents time points for a parameter as rows using variables such as AVISIT and ATPT. However, there may be cases when time points are needed as additional variables. For example, a tool or macro may require inputting time points as variables, or a statistician/reviewer may specifically request to see all the time points on the same row. This example demonstrates a way to support multiple analysis time points on one row and still maintain the ADaM principle of traceability.

In this example, a pharmacokinetics (PK) analysis is to be performed. Measurements of the concentration of study drugs are collected pre-dose and at a set of fixed times post-dose. Imputation for missing data is to be performed. To calculate the Area Under the Curve (AUC), the statistician provided code that used time points as variables. For facilitating visual review of the data and creating listings, it was requested to present the data in a horizontal structure.

DATA FLOW

The approach demonstrated is to first create a BDS dataset, making use of traceability built into the BDS standard to explain the origin, derivation, imputation, and any other complexity behind each analysis value. Within the BDS dataset, two variables, ATPT and custom variable ATPTCD are created with the intention of using their values as the variable label and name in a subsequent horizontal structure dataset. After the BDS dataset is created it is transposed into a horizontal dataset which is used to support statistical analysis and review.

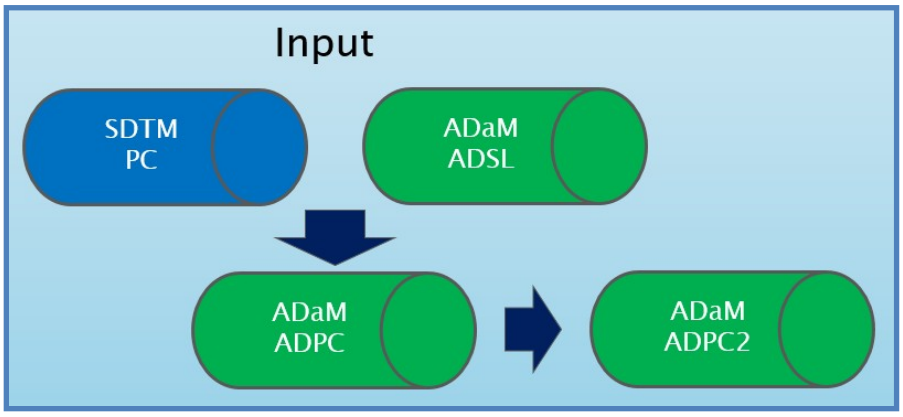


Figure 5: Example of Multiple Time Points on Same Row Traceability

TRACEABLITY NEEDS

Table 32 shows the variable metadata for ADPC. This dataset contains traceability variables to the SDTM PC dataset for records copied from PC, and derived records for time points which are missing in SDTM. The variable ATPTCD is created with the intent to transpose this dataset.

Note that only variables used to illustrate the concept of traceability are shown in the tables.

Table 32: Variable Metadata for ADPC

Variable Name	Where Condition	Variable Metadata
STUDYID		PC.STUDYID
USUBJID		PC.USUBJID

Variable Name	Where Condition	Variable Metadata
TRTP		ADSL.TRT01P
ATPT		Proper case of PC.PCTEST concatenated with PC.PCTPT Create new records for any subjects without a full set of time points: predose, 1/2/4/8/16/24/48 hours post dose
ATPTCD		One-to-one mapping with ATPT, map value "Drug A" to "DA_", map value "Drug B" to "DB_", and map time point "Predose" to "PD", map time point "X Hours Postdose" to "X"ex: "Drug A 48 Hours Postdose" is mapped to "DA_48", "Drug B Predose" is mapped to "DB_PD"
PARAMCD		PC.PCTESTCD
PARAM		PC.PCTEST (PC.PCSTRESU)
DTYPE		Set to "ASSIGNED" if PC.PCSTRESN is missing and time point is pre-dose. Set to "INTERP" if PC.PCSTRESN is missing and time point is 2 through 48 hours postdose, and subject has at least 4 of those 6 time points populated Otherwise set to missing (time point 1 hour postdose will not be imputed)
AVAL	DTYPE EQ ' '	PC.PCSTRESN
	DTYPE EQ 'ASSIGNED'	Assign value 0
	DTYPE EQ 'INTERP'	Regress existing values for the subject from time points 2 through 48 hours post-dose using exponential decay function, using 2 hours post-dose as T(0), impute missing value using extrapolated curve and the record time point. Please see Analysis Data Reviewer's Guide (ADRG) section 2.3.0 for details
PCSEQ		PC.PCSEQ
PCTEST		PC.PCTEST
PCTPT		PC.PCTPT

Table 33 shows the dataset metadata for ADPC2. It describes that ADPC2 as being a transposed version of ADPC.

Table 33: Dataset Metadata for ADPC2

Dataset	Description	Class	Structure	Description
ADPC2	PK Analysis 2	BASIC DATA STRUCTURE	One record per subject per parameter	This dataset is derived from ADPC by transposing AVAL by USUBJID, using the values of ATPTCD as new variable names and the values of ATPT as new variable labels.

Table 34 shows the variable metadata for ADPC2. The metadata for transposed variables is straightforward, and for the complex derivation of the AUC the reviewer is directed to the ADRG.

Table 34: Variable Metadata for ADPC2

Variable Name	Variable Label	Variable Metadata
STUDYID	Study Identifier	ADPC.STUDYID
USUBJID	Unique Subject Identifier	ADPC.USUBJID
TRTP	Planned Treatment	ADPC.TRTP
PARAMCD	Parameter Code	Assign value 'AUC'
PARAM	Parameter	Assign value 'Area Under the Curve'

Variable Name	Variable Label	Variable Metadata
AVAL	Analysis Value	Please see Analysis Data Reviewer's Guide (ADRG) section 2.3.1 for the statistical formula and SAS code used to calculate AUC
DA_PD	Drug A Predose	ADPC.AVAL where ATPTCD='DA_PD'
DA_1	Drug A 1 Hour Postdose	ADPC.AVAL where ATPTCD='DA_1'
DA_2	Drug A 2 Hours Postdose	ADPC.AVAL where ATPTCD='DA_2'
DA_4	Drug A 4 Hours Postdose	ADPC.AVAL where ATPTCD='DA_4'
DA_8	Drug A 8 Hours Postdose	ADPC.AVAL where ATPTCD='DA_8'
DA_16	Drug A 16 Hours Postdose	ADPC.AVAL where ATPTCD='DA_16'
DA_24	Drug A 24 Hours Postdose	ADPC.AVAL where ATPTCD='DA_24'
DA_48	Drug A 48 Hours Postdose	ADPC.AVAL where ATPTCD='DA_48'
DA_PD	Drug B Predose	ADPC.AVAL where ATPTCD='DB_PD'
DB_1	Drug A 1 Hour Postdose	ADPC.AVAL where ATPTCD='DB_1'
DB_2	Drug A 2 Hours Postdose	ADPC.AVAL where ATPTCD='DB_2'
DB_4	Drug A 4 Hours Postdose	ADPC.AVAL where ATPTCD='DB_4'
DB_8	Drug A 8 Hours Postdose	ADPC.AVAL where ATPTCD='DB_8'
DB_16	Drug A 16 Hours Postdose	ADPC.AVAL where ATPTCD='DB_16'
DB_24	Drug A 24 Hours Postdose	ADPC.AVAL where ATPTCD='DB_24'
DB_48	Drug A 48 Hours Postdose	ADPC.AVAL where ATPTCD='DB_48'

INPUT AND ANALYSIS DATA

The dataset below shows a sample implementation of the ADPC, ADPC2 metadata above. In the ADPC dataset shown in **Table 35**, it is clear which records are from SDTM, and which records are missing and imputed by examining the PCSEQ and DTYPE variables.

Please note only a small subset of variables are kept in this example for presentation purposes.

Table 35: ADPC Sample Records

Row	STUDYID	USUBJID	TRTP	ATPT	ATPTCD	PARAM
1	XYZ	XYZ-001	DRUG A 200 MG	Drug A Predose	DA_PD	DRUG A (ng/mL)
2	XYZ	XYZ-001	DRUG A 200 MG	Drug A 1 Hour Postdose	DA_1	DRUG A (ng/mL)
3	XYZ	XYZ-001	DRUG A 200 MG	Drug A 2 Hours Postdose	DA_2	DRUG A (ng/mL)
4	XYZ	XYZ-001	DRUG A 200 MG	Drug A 4 Hours Postdose	DA_4	DRUG A (ng/mL)
5	XYZ	XYZ-001	DRUG A 200 MG	Drug A 8 Hours Postdose	DA_8	DRUG A (ng/mL)
6	XYZ	XYZ-001	DRUG A 200 MG	Drug A 16 Hours Postdose	DA_16	DRUG A (ng/mL)
7	XYZ	XYZ-001	DRUG A 200 MG	Drug A 24 Hours Postdose	DA_24	DRUG A (ng/mL)
8	XYZ	XYZ-001	DRUG A 200 MG	Drug A 48 Hours Postdose	DA_48	DRUG A (ng/mL)

Row	PARAMCD	DTYPE	AVAL	PCSEQ	PCTEST	PCTPT
1 (cont)	DRUGA		0	1	DRUG A	PREDOSE
2 (cont)	DRUGA		169	2	DRUG A	1 HOUR POSTDOSE

Row	PARAMCD	DTYPE	AVAL	PCSEQ	PCTEST	PCTPT
3 (cont)	DRUGA		233	3	DRUG A	2 HOURS POSTDOSE
4 (cont)	DRUGA		170	4	DRUG A	4 HOURS POSTDOSE
5 (cont)	DRUGA		132	5	DRUG A	8 HOURS POSTDOSE
6 (cont)	DRUGA		70	6	DRUG A	16 HOURS POSTDOSE
7 (cont)	DRUGA	INTERP	32			
8 (cont)	DRUGA	INTERP	4			

Table 36: ADPC2 Sample Records

Row	STUDYID	USUBJID	TRTP	PARAMCD	PARAM	AVAL	DA_PD	DA_1
	Study Identifier	Unique Subject Identifier	Planned Treatment	Parameter Code	Parameter	Analysis Value	Drug A Predose	Drug A 1 Hour Postdose
1	XYZ	XYZ-001	DRUG A 200 MG	AUC	Area Under the Curve	2730	0	169
2	XYZ	XYZ-002	DRUG A 100 MG + DRUG B 50 MG	AUC	Area Under the Curve	3328	0	82
3	XYZ	XYZ-003	DRUG A 200 MG	AUC	Area Under the Curve	3003	0	182
4	XYZ	XYZ-004	DRUG A 100 MG + DRUG B 50 MG	AUC	Area Under the Curve	3168	0	
5	XYZ	XYZ-005	DRUG A 100 MG + DRUG B 50 MG	AUC	Area Under the Curve		0	90
6	XYZ	XYZ-006	DRUG A 200 MG	AUC	Area Under the Curve	2751	0	160

Row	DA_2	DA_4	DA_8	DA_16	DA_24	DA_48
	Drug A 2 Hours Postdose	Drug A 4 Hours Postdose	Drug A 8 Hours Postdose	Drug A 16 Hours Postdose	Drug A 24 Hours Postdose	Drug A 48 Hours Postdose
1 (cont)	233	170	132	70	32	4
2 (cont)	125	117	104	81	65	32
3 (cont)	255	191	145	79	37	7
4 (cont)	121	111	99	77	61	29
5 (cont)	141					
6 (cont)	235	175	132	71	34	5

Please note if creating the ADPC2 dataset, **Table 36** by itself without ADPC, it would not be clear which time point values are collected and which are imputed, so neither data point traceability nor derivation support would be present. By having both ADPC and ADPC2, it is possible to review the AUC derivation in ADPC2, the imputation for missing values in ADPC, and data point traceability between these two datasets and the SDTM PC dataset is clear.

OTHER USES

This example demonstrated a horizontal dataset with the time points of analysis variables reorganized as variables. Each data value can still be traced back across derivations to its SDTM source using variable metadata, and data point traceability provided by the BDS standard.

The final dataset ADPC2 is a BDS class dataset due to the addition of variables AVAL, PARAMCD, and PARAM, which are necessitated by the computation of AUC. If the creation of a parameter was not needed, and simply transposing ADPC would have supported tables and figures, then the final dataset could remain as an ADaM - Other class dataset (named ADPCT as an example).

CONCLUSION

Five examples were shown to demonstrate how data, metadata, and even intermediate datasets, can all provide traceability when creating ADaM datasets. Each of these examples comes from the ADaM Traceability Examples document now in development.

When deciding how to create ADaM datasets, the authors encourage you to ask yourself the following questions:

- Can the end-user determine which data is copied from SDTM and which is derived?
- Can the end-user determine how each variable and row was created in the dataset?
- Can the end-user trace back to the SDTM data that was used to create the value used for analysis?

By considering the perspective of the end-user, traceability can be built in a natural and useful way.

ACKNOWLEDGEMENTS

We would like to thank Sandra Minjoe and Nancy Brucken for their support and review of the content in this paper. We would like to thank the entire ADaM Traceability Team for authoring, reviewing, and revising the examples presented in this paper.

REFERENCES

All CDISC documents referenced in this paper can be downloaded from https://www.cdisc.org/.

RECOMMENDED READING

- Analysis Data Model Implementation Guide version 1.1
- Analysis Data Model (ADaM) Examples in Commonly Used Statistical Analysis Methods
- CDISC Define-XML Specification Version 2.0

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Wayne Zhong Accretion Softworks Wayne321@gmail.com	**Richann Watson** DataRich Consulting richann.watson@datarichconsulting.com
Daphne Ewing CSL Behring Daphne.Ewing@cslbehring.com	**Jasmine Zhang** Boehringer Ingelheim zhangjingwendm@gmail.com